GenCore Version 4.5
Copyright (c) 1993 2000 Compugen Ltd.

tion search, using sw model

November 19, 2001, 16:59:44 : Search time 1262.29 Seconds
                                            (without alignments)
                                            7683.084 Million cell updates/sec

title: US-09-516-052-1
Perfect Score: 627
Sequence: 1 dataaraactcattatatt.........attatgtcaatatataatga t 27

Scoring table: IDENTITY_Nuc
                 Gapop 10.0 , Gapext 1.0
Searched: 44457 seqs, 773487588 residues

Total number of hits satisfying chosen parameters:
Maximum DB seq length: 2688414
Minimum DB seq length:

Post processing: Minimum Match 0%
                  Maximum Match 100%
                  List first 45 summaries

Database:
                  Sequential *

1:	db_bal: *
2:	db_fst2: *
3:	db_bst: *
4:	db_fol: *
5:	db_in2: *
6:	db_in1: *
7:	db_cm: *
8:	db_ew: *
9:	db_pat1: *
10:	db_pat2: *
11:	db_phr: *
12:	db_pl2: *
13:	db_pl3: *
14:	db_fol: *
15:	db_fol4: *
16:	db_u2: *
17:	db_ba1: *
18:	db_ba2: *
19:	db_hrqe: *
20:	db_hrqo_liv: *
21:	db_hrqo_rol: *
22:	db_hrq_pum1: *
23:	db_hrq_pum1: *
24:	db_hrq_pum2: *
25:	db_hrq_pum3: *
26:	db_hrq_pum4: *
27:	db_hrq_pum5: *
28:	db_hrq_pum6: *
29:	db_hrq_pum7: *
30:	db_hrq_pum8: *
31:	em_bta1: *
32:	em_bta2: *
33:	em_bta3: *
34:	em_bta4: *
35:	em_hum1: *
36:	em_hum2: *
37:	em_hum3: *
38:	em_hum4: *
39:	em_hom7: *
40:	em_hum7: *
41:	em_inv: *
42:	em_mus: *
43:	em_or1: *

44:	em_ov: *
45:	em_pat: *
46:	em_ph: *
47:	em_pl: *
48:	em_ro: *
49:	em_sts: *
50:	em_sy: *
51:	em_sy: *
52:	em_vi: *
53:	db_stst1: *
54:	db_stst2: *
55:	db_stst3: *
56:	db_sy: *
57:	db_uu: *
58:	db_vil: *
59:	db_vi2: *
60:	db_htg1: *
61:	db_htg2: *
62:	db_htg3: *
63:	db_htg4: *
64:	db_htg5: *
65:	db_htg6: *
66:	db_htg7: *
67:	db_htg8: *
68:	db_htg9: *
69:	db_htg10: *
70:	db_htg11: *
71:	db_htg12: *
72:	db_htg13: *
73:	db_htg14: *
74:	db_htg15: *
75:	db_htg16: *
76:	db_htg17: *
77:	db_htg18: *
78:	db_htg19: *
79:	db_htg20: *
80:	db_htg21: *
81:	db_htg22: *
82:	db_htg23: *
83:	db_htg24: *
84:	db_htg25: *
85:	db_pr1: *
86:	db_pr2: *
87:	db_pr3: *
88:	db_pr4: *
89:	db_pr5: *
90:	db_pr6: *
91:	db_pr7: *
92:	db_pr8: *
93:	db_pr9: *
94:	db_ro1: *
95:	db_ro2: *
96:	db_in4: *
97:	db_pr10: *
98:	em_ba3: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	627	100.0	826	12	AF036684	Arabidops
2	625.4	99.7	82875	12	AF014482	Genomic s
3	174.4	27.8	90117	12	AB025628	Arabidops
4	154.2	21.4	874	1	ATUAU4H	Arabidopsis
5	134	21.4	870	15	X54714	Z.mays mRNA
6	133.2	21.2	197976	1	ATUHRIV34	Arabidopsis
7	133.2	21.2	206606	13	APUYA1	Arabidopsis
8	131.4	21.0	141808	83	AP003266	Oryza Sat

KEYWORDS   .
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 82875)
  AUTHORS   Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
            Khan,S., Kim,C.J., Altafi,H., Bot,Q., Chin,C., Chiou,C., Choi,E.,
            Chin,J., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
            Lam,B., Lee,J., Lenz,C., Liu,J., Liu,A., Liu,S.,
            Miklasky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwarz,T.,
            Southwick,A., Thai,A., Toriumi,M., Vaysberg,M., Ya,C.,
            Federspiel,N., Theologis,A. and Ecker,J.R.
  TITLE   Genomic sequence for Arabidopsis thaliana BAC T26F17 from
            chromosome 1
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 82875)
  AUTHORS   Ecker,J.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-NOV-1999) Arabidopsis Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3 (bases 1 to 82875)
  AUTHORS   Ecker,J.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-NOV-1999) Arabidopsis Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4 (bases 1 to 82875)
  AUTHORS   Buehler,E., Chao,Q., Johnson-Hopson,C.,
            Shinn,P., Brooks,S., Kim,C.J., Altafi,H., Chin,C., Choi,E.,
            Chin,J., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
            Lam,B., Lee,J., Lenz,C., Liu,J., Liu,A., Liu,S.,
            Miklasky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwarz,T.,
            Southwick,A., Thai,A., Toriumi,M., Vaysberg,M., Ya,C., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Arabidopsis Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE   5 (bases 1 to 82875)
  AUTHORS   Buehler,E., Chao,Q.,
            Shinn,P., Brooks,S., Kim,C.J., Altafi,H., Chin,C.,
            Chin,J., Conway,A., Gonzales,A., Hansen,N.,
            Lam,B., Lee,J., Lenz,C., Liu,J., Liu,A., Liu,S.,
            Miklasky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwarz,T.,
            Southwick,A., Thai,A., Toriumi,M., Vaysberg,M., Ya,C., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Arabidopsis Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT   This sequence version replaced gi:6429478.
FEATURES   Location/Qualifiers
     source   1..82875
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="T26F17"
     CDS   398,676. .953,145. .1474)
                     /codon_start=1
                     /note="unknown; similar to EST gb:H74586"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="T26F17.1"
                     /protein_id="AAF16521.1"
                     /db_xref="gi:6652572H"

     tRNA   complement(1904. .1974)
                     /note="codon recognized: CAA"
                     /product="tRNA-Gln"
     CDS   complement(join(2461. .2747,2841. .4249,424. .4420))
                     /codon_start=1
                     /note="hypothetical protein"
                     /evidence=not_experimental
                     /product="T26F17.2"
                     /protein_id="AAF16547.1"
                     /db_xref="gi:6652748"
                     /translation="MMHCHFLLLRTVWKSSGHLVPFLHAKVAAPIAVA-WPKAENEVVEL
                     KQKLEDMAIRNIVLEDRVSHLLQSALKEIVRQLRQFRQEKNIQAAVTESIKILHSAN
                     ICLEKRVLLLGKLAEZAASHNMMLRRENLSHSSHNLTWLEMLSSY-AAFSAKSDKDH
                     1SKIAKLEAETPPTILARTSSESSNQSVQSHSKRKSVSKSVSWASSSAEVSE..."
     CDS   complement(7212. .7601)
                     /note="unknown; similar to EST gb:AA549241"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="T26F17.3"
                     /protein_id="AAF16546.1"
                     /db_xref="gi:6652747"
     CDS   complement(8023. .8304)
                     /note="unknown; similar to ESTs gb:H21171 and
                     gb:AI992884.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="T26F17.4"
                     /protein_id="AAF16545.1"
                     /db_xref="gi:6652746"
     CDS   complement(join(11586. .11828,11940. .12640. .13446,
                     14642. .14884,15072. .15171,15188,15658. .15975. .16094,
                     16170. .16438))
                     /note="similar to pollen specific protein precursor
                     emb:CAA24505; similar to ESTs emb:Y14024, gb:AI949104.1,
                     gb:R40156, gb:AI774044.1, gb:AI724078.1, gb:AI725799, and
                     gb:AI992516.1"
                     /codon_start=1
                     /translation="MLVRSKVAL..."
                     /evidence=not_experimental
                     /product="T26F17.6"
                     /protein_id="AAF16544.1"
                     /db_xref="gi:6652745"

/organism "Arabidopsis thaliana"
/strain "Columbia"
/db_xref "taxon:3702"

/clone_lib "Mtsu1 P1"
/note "45a7"
/clone_lib "Mtsu1 P1"

/note "gis is reported in Acc# AB018117"

/product "DNA binding protein like"

/note "gene_id:MNJ7.2"
/codon_start 1

/product "DNA binding protein like"

/note "gene_id:MNJ7.1"
/codon_start 1

/note "contains similarity to unknown protein"
/codon_start 1
/protein_id "BAB09069.1"
/db_xref "GI:9758771"

/product "unknown protein"
/note "gene_id:MNJ7.7"

/product "selenium binding protein like"
/protein_id "BAB09072.1"
/db_xref "GI:9758774"

/product "modulin like protein"

/note "gene_id:MNJ7.6"
/codon_start 1

/product "modulin like protein"
/protein_id "BAB09073.1"
/db_xref "GI:9758775"

/note "gene_id:MNJ7.7"

/product "unknown protein"
/protein_id "BAB09074.1"
/db_xref "GI:9758776"

/note "contains similarity to unknown protein"
/gene_id:MNJ7.8
/codon_start 1
/evidence not_experimental

/gene="At4g14420"
/note="strong similarity to Nicotiana tabacum ORF able to
induce HR-like lesions, PATX:g17624945,
contains EST db:AI999161.1"
/codon_start 1
/product="elicitor like protein"
/protein_id "AAB78484.1"
/db_xref "GI:7268148"
/translation "MEA..."

/gene="At4g14420"
/number-1
complement(5777..5864)
/gene="At4g14420"

/gene="At4g14420"
complement(5865..5963)
/number-2
/gene="At4g14420"

/gene="At4g14420"
complement(5964..6050)
/number-2
/gene="At4g14420"

/gene="At4g14420"
complement(6051..6140)
/number-3
/gene="At4g14420"

/gene="At4g14420"
complement(6141..6247)
/number-3
/gene="At4g14420"

/gene="At4g14420"
complement(6248..6352)
/number-4
/gene="At4g14420"

/gene="At4g14420"
complement(6353..7167)
/number-4
/gene="At4g14420"

complement(6454..7167)
/number-5
/gene="At4g14420"

complement(7168..7245)
/number-5
/gene="At4g14420"

complement(8407..9129)
/gene="At4g14430"
/number-1

8407..9129
/gene="At4g14430"

complement(8407..9129)
/gene="At4g14430"

complement(8407..9129)
/gene="At4g14430"
/note "similarity to carnitine racemase, Escherichia coli
contains EST gb:H36845, AI994747.1, 17468"
/codon_start 1
/product "carnitine racemase like protein"
/protein_id "CAB78485.1"
/db_xref "GI:7268149"
/translation "MVTLKRGLLLLEL..."

complement(10443..11159)
/gene="At4g14440"

complement(10443..11159)
/gene="At4g14440"
/note "similarity to carnitine racemase, Escherichia coli
contains EST gb:AI994408.1, AA494852, H46791, AA494904,
T76585, H77091, 149285, AA597488, Z17680"
/codon_start 1
/product "carnitine racemase like protein"
/protein_id "CAB78486.1"
/db_xref "GI:7268150"
/translation "MEIFKRGLLLLELLLL..."

complement(10443..11159)
/number 1
/gene="At4g14440"

10443..11159
/gene="At4g14440"

complement(10443..11159)
/gene="At4g14440"

BASE COUNT
ORIGIN

QUERY Match
Best Local Similarity
Matches

RESULT
AT_THRIV44
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
source

gene

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

exon

exon

gene

gene

gene

exon

gene

annotation, at this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at http://www...
this transcript has an overlap with AtPCAO at the 3' end and an
overlap with AtPCA2 at the 5' end and...

FEATURES

source   1..26806
            /organism="Arabidopsis thaliana"
            /Bt_xref "taxon:3702"
            /variety "Columbia"

source   1..26121
            /organism="Arabidopsis thaliana"
            /Bt_xref "taxon:3702"
            /chromosome="4"

gene   1806
            /gene "dl3205c"

mRNA   complement(645..744)
            /gene "dl3205c"

gene   complement(645..744)
            /gene "dl3205c"

CDS   complement(join(645..744,829..905,1054..1222,1931..1806))
            /number 1
            /gene "dl3205c"   744,829, 905,1054, 1222,1931, 1806))
            /product "ribosomal protein"
            /protein_id "AR10211.1"
            /db_xref "GI:2247780"
            /note "SWISS PROT:O23290"
            ...

intron   codon_start 1
            strong similarity to ribosomal protein 141 Candida
            maltosa
            complement(1745..828)
            /number 1

exon   complement(829..904)
            /number 2

intron   complement(905..1052)
            /number 2

exon   complement(1053..1222)
            /number 3

intron   complement(1223..1806)
            /number 3

exon   complement(1804..1806)
            /number 4

gene   complement to kinesin osm-4 (see also list)
            contains est
            elegans
            contains est gb:496806

gene   /gene "dl3205w"
            /number 1

exon   2234..2484
            /number 1
            /gene "dl3205w"
            /product "kinesin like protein"
            /protein_id "AR10212.1"
            /db_xref "GI:2247790"

CDS   join(2234..2484,4863,4861, 5268,5278, 468,
            4564, 4947,5000, 5212,6540, 6428,6469, 6560,6708, 6749,
            7516, 7784)

[right column continued]

intron   /gene="dl3205w"
            /number 2
2914..3069 2914.. 3069
            /number 2

intron   4070.. 4160
            /gene "dl3205w"
            /number 3

exon   4161.. 3268
            /gene "dl3205w"
            /number 2

intron   3269.. 3457
            /gene "dl3205w"
            /number 3

exon   3458.. 3468
            /gene "dl3205w"
            /number 3

intron   3469.. 3568
            /gene "dl3205w"
            /number 4

exon   5213.. 6339
            /gene "dl3205w"
            /number 6

intron   6340.. 6428
            /gene "dl3205w"
            /number 6

exon   6429.. 6468
            /gene "dl3205w"
            /number 7

intron   6469.. 6500
            /gene "dl3205w"
            /number 7

exon   4958.. 4999
            /gene "dl3205w"
            /number 5

intron   5000.. 5212
            /gene "dl3205w"
            /number 5

...   7559.. 4957
            /gene "dl3205w"
            /number 5

gene   9669
            /gene "dl3210c"

complement(join(6513.. 6772,6898, 7624,7289, 7752,
7451.. 7524,7758, 7784,7870, 7954,8045, 8109,9216, 8277,
8489.. 8559,8688. 8788,8860. 9008,9099, 9168,9444, 9475,
9576.. 9669))
            /gene "dl3210c"

gene   /gene "Skil mRNA for casein kinase I
            contains EST gb:957978;1054381"

CDS   complement(join(6513.. 6772,6898, 7624,7289, 7752,
            7451.. 7524,7758, 7794,7870, 7954,8045, 8109,9216, 8277,
            8489.. 8559,8688. 8788,8860. 9008,9099, 9168,9444, 9475,
            9576.. 9669))
            /note "Skil mRNA for casein kinase I
            contains EST gb:957978;1054381"
            /product "casein kinase"
            /protein_id "AR10213.1"
            /db_xref "GI:2247791"
            /note "SWISS PROT:O23290"
            /translation "MDWNGKMHV..."

Left column

COMMENT Agricultural Resources, Rice Genome Research Program: Kannondai
         2-1-2 Tsukuba, Ibaraki 305 8602, Japan
         (E-mail: ... Fax:81-298-48-7468)
         Total: 46-7441, Fax:81-298-48-7468)

         NOTE: It currently consists of 1 contigs. Gaps between the contigs
         are represented as runs of N. The order of the pieces is believed
         to be correct as given, however the sizes of the gaps between them
         are based on estimates that have been provided by the submitter. This
         sequence will be replaced by the finished sequence as soon as it is
         available and the accession number will be preserved.
         * NOTE: this is a 'working draft' sequence.
         * This sequence will be replaced
         * by the finished sequence as soon as it is available and
         * the accession number will be preserved.

FEATURES   Location/Qualifiers
     source   1..174685
                     /organism="Oryza sativa"
                     /cultivar=...
                     /db_xref="taxon:4530"
                     /chromosome="1"

BASE COUNT   ...
ORIGIN

Query Match   21.0%; Score 131.4; DB 84; Length 174886;
Best Local Similarity   65.5%;
Matches 142 Conservative ... Mismatches 101; Indels 0; Gaps 0;

REFERENCE
AUTHORS
TITLE
JOURNAL

ACCESSION   AP005409/C
LOCUS
DEFINITION Arabidopsis thaliana chromosome 11 section 254 of 255 of the
            complete sequence. Sequence from clones T30B22, T17A22, T9J24.
VERSION   AP005409.2 GI:6654945
KEYWORDS   HTG.
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 107377)
AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Vamathevan,J.,
            Fujii,C.Y., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
TITLE   Sequence and analysis of chromosome 2 of the plant Arabidopsis

Right column

JOURNAL   thaliana
          Nature 402 (6763), 761-768 (1999)
PUBMED   10617197
REFERENCE 2 (bases 1 to 107377)
AUTHORS   Lin,X.
TITLE   Direct Submission
JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
          Medical Center Dr., Rockville, MD 20850, USA
COMMENT   On Dec 17, 1999 this sequence version replaced gi:6748275.
          The sequence and annotation of chromosome 2 were merged from those
          of the individual clones on this chromosome after removing
          overlaps. For detailed information, please see the TIGR web site
          (http://www.tigr.org/tdb/at/at.html).

          Genes were identified by a combination of three methods: Gene
          prediction programs including GRAIL
          (ftp://... Oak Ridge National ...), Genefinder (Phil Green,
          University of Washington), Genscan (Chris Burge,
          (http://genome.stanford.edu/GENSCANW.html), and Netplantgene
          (http://www.cbs.dtu.dk/services/NetPGene/). Searches of the
          complete sequence against a peptide database and plant EST
          databases at TIGR, and manual curations based on these analyses.
          Annotated genes are named to indicate the level of evidence for
          their annotation. Genes with similarity to other proteins are named
          after the database hits. Genes without significant peptide
          similarity but with EST similarity are named as 'unknown' proteins.
          Genes without protein or EST similarity, that are predicted by two
          or more gene prediction programs over most of their length are
          annotated as 'hypothetical' proteins. Genes encoding tRNAs are
          predicted by tRNAscan-SE (Sean Eddy,
          (http://...), or FRNA. Simple repeats were
          identified by repeatmasker (Arian Smit,
          (http://...). Genes are
          numbered from the top to bottom of the chromosome.

          Address all correspondence to: at@tigr.org.

FEATURES   Location/Qualifiers
     source   1..107377
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     misc_feature   complement(<1..821)
                     /note="Sequence from clone T30B22"
     mRNA   join(<291..>2508)
                     /gene="At2g47620"
     gene   <291..>2508
                     /gene="At2g47620"
                     /note="T30B22.7"
     CDS   join(291..348,504..1024,1120..1549,2112..2243,2342..2508)
                     /gene="At2g47620"
                     /codon_start=1
                     /product="putative SWI/SNF family transcription activator"
                     /protein_id="AAG13618.2"
                     /db_xref="GI:6598454"
                     /translation="MEATPSAEIEIYTPAQSSWFDWDDIHEJERREFARFFTSSI
                     TRPPKVRKEYRDPIINKREPDYCRRLPFVSKEFVEDVNLLGRVEFIFLKKWDLINFS
                     SSLKKNDHLLSVEMAKIEQQTEEKKYEANTSLREFVEDVNLLGRVEFIFLKKWDLINFS
                     YSDVFSDLKKFEAVEVCAHEGES JUST TKHKGIVHICHK TKRCNCHRHLADHFK"

REFERENCE
  AUTHORS   Eukarya; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
  REFERENCE 1 (bases 1 to 82403)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE   Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20049487
   PUBMED   10617197
  REFERENCE 2 (bases 1 to 82403)
  AUTHORS   Lin,X.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA

FEATURES   Location/Qualifiers
     source   1..82403
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
     misc_feature   1..82202
                     /note="sequence from clone T10P5"
     repeat_region   1197..1306
                     /rpt_family="(TTAAAA)n"
     repeat_region   complement(4226..4308)
                     /rpt_family="(TTAAAA)n"

     repeat_region   7744..7865
                     /rpt_family="(TTAAAA)n"
     repeat_region   complement(7498..11670)
                     /rpt_family="(TTAAAA)n"
     gene   complement(7498..11670)
                     /gene="At2g13470"
                     /pseudo
     gene   complement(12098..12827)
                     /gene="At2g13480"
                     /note="T10P5.1; similar to GB:AAA75525.4"
                     /pseudo
     repeat_region   complement(13418..13471)
                     /rpt_family="(TTAAAA)n"
     gene   14215..18084
                     /gene="At2g13490"
                     /note="T10P5.3"
     repeat_region   complement(18097..18209)
                     /rpt_family="(CAAAA)n"
     mRNA   join(<18186..19150,19298..>19526)
                     /gene="At2g13500"
     gene   <18186..>19526
                     /gene="At2g13500"
                     /note="T10P5.4; similar to GB:AAA75254; some members of
                     this protein family have a weak CCH zinc fingers that is
                     mostly from retroviral gag proteins (nucleocapsid)"
     CDS   join(18186..19150,19298..19526)
                     /gene="At2g13500"
                     /codon_start=1
                     /product="putative Gag-like non-ltr retroelement protein"
                     /protein_id="AAD22675.1"
                     /db_xref="GI:4558657"
                     /translation="MVDVWDEIVQRIEDSCINITLFIPHEAYWMAVEVFNRLSLIAPEC
                     LFKVDEHAVINALEKAWCLIAHVHGKGIAIVVQLLFGTTIIILSVGPFEYWLFNRWTRWM
                     FVASQRWEAAFSLHEVTTIDLRWVQMRCIPIIYVPEKTPLEIAQEHLGEIISLDEHDAIS
                     TQIAYIRVRVREGITDRLRFFVFQRITFESGETALIRFQYEPRLRKIYSFPRFTININYPP
                     YRQKFQSIIRERAVFRDSVLRSSHTSQMTDSSIQVPLHPPPVPSSDAIPSGIMK
                     VEFVGESSKGVGZGLSHRHHLCLSSKREHLVILQPHWSPQHKQDTPIQHMRSLAP
                     PKKR"
     gene   complement(<20428..>21544)
                     /gene="At2g13510"
                     /pseudo
     gene   complement(<20428..>21544)
                     /gene="At2g13510"
                     /note="T10P5.5; similar to GB:AAA75254; some members of
                     this protein family have a weak CCH zinc fingers that is
                     mostly from retroviral gag proteins (nucleocapsid)"
     mRNA   complement(join(20428..20746,20999..>21544))
                     /gene="At2g13510"
     CDS   complement(join(20428..20746,20999..21544))
                     /gene="At2g13510"
                     /codon_start=1
                     /product="putative Gag-like non-ltr retroelement protein"
                     /protein_id="AAD22676.1"
                     /db_xref="GI:4558658"
                     /translation="MAIKIWHEIAQEEIGIEGDHVTLEIPHEAYWMAVEVFNRLSLIAPE
                     NPHVQNLYSVIAALPKSWGLTTHVHGKVIIDAIVQPLPHIEMDIVSVGKREYWLEFNNR
                     VASQKWTAFAENIVTTIDLLVQMRCIIIJYVFHTALFIAQFFIFIISLDEHDAIS
                     TQIAYIRVRVQIGITERLRVHAPILNEHELAGASIYFPSSKNAAHQIYMVFSTQQRVG
                     RHHVFFPSSNINPSSIAALIESGVKRVLVGFSIRPEPIAJLSRNRRRIGHSSKREHLVIL
                     QSIDNEKAN"
     gene   complement(21749..25814)
                     /gene="At2g13520"
                     /note="T10P5.6"
     repeat_region   21554..21667
                     /rpt_family="(TTAAAA)n"
     mRNA   join(<27459..27581,27643..27823,27827..27849,28249..28449,
                     28529..28591,28765..28942,29126..29254,29478..29544,
                     29724..29923..30116,30294,30756..30904,30971,
                     31063..31181,31489..31627,31829..32329,32420..32770,
                     33154..33472)

```
VERSION     AC006424.1  GI:6042674
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 110514)
  AUTHORS   Osborne,B.I., Vysotskaia,V.S., Toriumi,M., Yu,G.-L., Oji,Q.,
            Buehler,E., Conway,A.R., Sun,H., Dewar,K., Feng,J., Kim,C.,
            Kut,C.-J., Liu,Y., Shian,P., Sun,H., Davis,R.W., Eske,J.R.,
            Federspiel,N.A. and Theologis,A.
  TITLE     The Sequence of the BAC F7G19 from Arabidopsis thaliana chromosome 1
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   3  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   4  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-Feb-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   5  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   6  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   7  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   8  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
REFERENCE   9  (bases 1 to 110514)
  AUTHORS   Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1997), Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     On Aug 26, 1999 this sequence version replaced gi:1941998.
            The sequence of BAC F7G19 from Arabidopsis thaliana chromosome 1.
FEATURES             Location/Qualifiers
     source          1..110514
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
     gene            <1..2850
                     /gene="F7G19.1"
     CDS             join(971, 1274,1546, 1641,1688, 1817,2144, 2250,
                     2556, 2665,2740, 2850)
                     /gene="F7G19.1"
                     /note="similar to ATP-dependent Clp protease (db|D90915).
                     EST dbjN5461) comes from this gene."
                     /evidence=not_experimental
                     /codon_start=1
                     /protein_id="AAB70396.1"
```

```
                     /db_xref="GI:2342674"
                     /translation="MASLLASMNSLLFKSSSSHHTLPLFSSGIHH KITSYAFA
                     KASAKLPMPPINPKHPELSTIASIAANSPEKILNRIVANIVPPYLDIITSPOLMSSVA
                     QVERKSVAYNEHRPRTPPPDLPSMLLDPGKIVY LGMPTLDDEIFPTPQLVIAVTELVAAELM
                     YLGWLDPKPPIYIYINSTGTLRDGEIVAH VCVIAALSAGLKGVMMRPYYPMIAIVFA
                     KGEESEFDYFPLFI IPVMQLGVITRRGIL VPLLSKHTGNEVEFIVANVMRPTYMIAI VFA
                     KITPSYVIDRVSLGTVIF"
     tRNA            complement(3700..3771)
                     /note="codon recognized: GUC; Glu"
                     /product="tRNA-Glu"
     gene            4442..6513
                     /gene="F7G19.2"
     CDS             join(4442, 4646,4391, 4507,4747, 4854,4974, 5164,
                     5263, 5454,5551, 5669,5747, 5814,5916, 6216, 6446,
                     6446, 6513)
                     /gene="F7G19.2"
                     /note="similar to probable Mg dependent ATPase
                     (pir|S56671) FSTS db|T47842,db|AAA4798 comes from this
                     gene."
                     /evidence=not_experimental
                     /protein_id="AAB70397.1"
                     /db_xref="GI:2342675"
                     /translation="MAEFISFEGHGQLASMIIDDIGGHASHLLANEIRIKEESQKINID
                     LESVKEKIKIKENSEKIKINKQLIIVKPPITKRHIEHETAIFMHAHMIILSQKRKIV
                     VLKTSTROTIPLPVVLLVDPPTLPESHA V-VNKPSYLLI IIPPSPSPDSKKAMNVPEK
                     PFEDYNDIGGLEKOIQELVEAIVLPMTHHKFQFEKLGI RIPPKGVLLYGPPGTCKTLMAK
                     A/AAQTNATFIRFIAGPQLVGMFIGDEAKLVRDAFILLAKKRSVIIFIIDEIDAIGTKRP
                     DSEVSGDREVQRTMLEILNQLLDGESSDDRIKVIAATNRAIILDPALMRSRLDRKIEH
                     PHPTEEARGRLLQIIHSRKMNVNADVNFEELARSTDEFNAGIKAVVEAIMEAIRRQA
                     TEVNIEEDFNPETIGVAAKKFKASLNYYA"
     gene            complement(6693..9149)
                     /gene="F7G19.3"
     CDS             complement(join(6693, 6802,6892, 7491,7471, 7544,
                     7613, 7772,7805, 7793,8052, 8042,8795, 8844,8795,
                     9149))
                     /gene="F7G19.3"
                     /note="strong similarity to oryza NADPH oxidase
                     (db|X93301)"
                     /evidence=not_experimental
                     /codon_start=1
                     /protein_id="AAB70498.1"
                     /db_xref="GI:2342676"
                     /translation="MTSEIL IPTEPENVKEEAMNISYIEENWKGLWVLIIWLSIG"I
                     TLFTWKFPLGYFKKIVEFVEMGY-V VAKGSAICIH KIHNALILLPV SNEIIWEIKRSEL
                     IGSVVPPHDNINFHKVVAFGIAVGI H BIA-LSEIE-LIAAKNVPNFESE FGPDE
                     RPENYGWFMKGPPGWTGVTMVKIVAYVLMLXAYVLA-SWFEKNFANLFKSLKRLIDFNA WYS
                     HHLFVIVIVILLIVHGSFYVI.SREWYIHKITWMILAVIPVILLYAFFKRIIRAFHVCKAKVKV
                     LKVAVVPFNVISEVMSKRPKGFKYFSGQY IYIN-SDVSFLQWHFFSI ISASGDEX,SVH
                     YRNTEVLLIVGLSICAIPLIEILIEEVLRLYEKIERLISIEF-TPRWLFKNYVAFEKAY-TWV
                     IREQGSLEWFSRSRVMNNVVAEYYDSEGIMITFENYTSVTEEGIIARSALLIMASHBHARSG
                     IDIVSGTRVRTHFARFNWRSVFKIHVAYFHVRJAVGVIVCCNF-LLSSFELPLAQPSPF
                     ITIKFFEPHKENF"
     gene            complement(9410..10440)
                     /gene="F7G19.4"
     CDS             complement(join(9410, 9492,9582, 9640,9745, 9904,
                     10061, 10440))
                     /gene="F7G19.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F7G19.4"
                     /protein_id="AAB70499.1"
                     /db_xref="GI:2342677"
                     /translation="MESSEGET"NKISKCKAICSSINDEBYVEEITLEVRDFTINIMKA
                     KAILRSVLSGRIKFMVKSLSFASRKLDKSRSPFAMFAIRGLRFIAKNDAVGRNWDEVA
                     MEPIKLAVECKLEPSKSKEHCIEDMVESSEFVNEIFEALVFFPPTSSSSTKTELPETWE
                     QIFCNSEFDRLQIFFDMVDKNIDGRIIGDEVKFIIALSASANKLSKLNLNEVFTAAL
                     MEFLDRDNLGYIEVRCLLSNGYTVLISLLKIWLNF"
     gene            11553..13929
                     /gene="F7G19.5"
     CDS             join(11553, 11812,11908, 12119,12289, 13401,13304,
                     13419, 13929)
                     /gene="F7G19.5"
```

DEFINITION   Schistosoma mansoni nuclear factor Y transcription factor subunit B
             (NF-YB) mRNA, complete cds.
ACCESSION   AF327742
VERSION   AF327742.1
KEYWORDS   .
SOURCE   Schistosoma mansoni.
  ORGANISM   Schistosoma mansoni
             Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
             Trematoda; Digenea; Strigeidida; Schistosomatoidea;
             Schistosomatidae; Schistosoma.
REFERENCE   1 (bases 1 to 1547)
  AUTHORS   Serra,E., Zemzoumi,K., di Silvio,A., Mantovani,R., Lardans,V., and
             Dissous,C.
  TITLE   Conservation and divergence of NF-Y transcriptional activation
             in the trematode Schistosoma mansoni
  JOURNAL   Nucleic Acids Res. 26 (20), 3800-3805 (1998)
REFERENCE   2 (bases 1 to 1547)
  AUTHORS   Serra,E., Zemzoumi,K., Trolet,J. and Dissous,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-DEC-1997) INSERM U 167, Institut Pasteur, 1 rue du
             Professeur A. Calmette, Lille 59019, France
FEATURES   Location/Qualifiers
     source   1..1547
                     /organism="Schistosoma mansoni"
                     /db_xref="taxon:6183"
     CDS   53..529
                     /db_xref="GI:3132677"
                     /codon_start=1
                     /product="nuclear factor Y transcription factor subunit B"
                     /protein_id="AAC28780.1"
                     /translation="MVDTDNDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEF
                     ISFITSEASDKCQKEKRKTINGDDLLWAMATLGFEDYIEPLKVYLQKYREMEGDTKGS
                     AKGDDKGGSSSRGGDSSSGRQSLIMAVSHGGQGSNSHRNVIGSTAHGMSANMMSKRVPG
                     SKPQSNNYQPSGPSSLI"

BASE COUNT   449 a   401 c   248 g   449 t
ORIGIN

Query Match   73.9%;   Score 167.4;   DB 5; Length 1547;
Best Local Similarity   81.6%;   Pred. No. 3.4e-22;
Matches 170   Conservative   0; Mismatches 106;   Indels   0; Gaps   0;

Qy   81 tcatggtggacacagacaatgatcgattcttaccaataqctaatqtaaqtcqaatcat 140
            | |||||||| ||||| |||| |||||| ||||||||| ||||||||| |||||||
Db   180 tcatggtaagcatcctgttatagtttaatqtaatcttctaaaataattaaaaaaaatag 239

Qy   141 gaagaaggccctgcctgcaaatgctaaaataagtaaagatgcaaaaqaaacaqtacaa 200
            ||||||||| || |||| ||||||||| || | ||| || |||||| |||||||
Db   240 gaacaatggctatgtaaaataaaataaagaatgctaagaatgattataattat 289

Qy   201 qaatgtgtaagtgaattcataagttttattacaagtgaaqcaaqtgataaatgccaqaaa 260
            |||||||||||||| || |||| |||| ||||| ||||| || || |||| |||||
Db   300 gaatcaatttaggtttatttaaataatttagtaatatagtttaaaatgttaaaaa 359

Qy   261 gaaaaqaqaaaaacaataaatqqtqatqatctactqtqqqcaatqqcaacactaqqttt 320
            | |||||||| ||| || || | ||||| ||| || |||||| ||||| ||||
Db   360 gataatagttaattaacaaaattttgtaactataqtqqqcaatqqcaacactaqqtta 419

Qy   421 gggtcaagtagtagacgtggagattcaagtagtggacgtcaatctttg 456
            ||||||||||| || || ||| ||||| ||| ||||| ||| ||||
Db   44 tattaaccaaataatttctcatttaccaaataagttgattataattaatttccagaaat 455

Search complete: 1 November 14, 2001, 18:29:25
Add time: 47.91 sec